D280–D284 *Nucleic Acids Research, 2013, Vol. 41, Database issue* *Published online 17 November 2012*
doi:10.1093/nar/gks1061

Voronoia4RNA—a database of atomic packing densities of RNA structures and their complexes

Jochen Ismer[1], Alexander S. Rose[1], Johanna K. S. Tiemann[1], Andrean Goede[2], Kristian Rother[3] and Peter W. Hildebrand[1,*]

[1]Charité, Institute of Medical Physics and Biophysics, Proteinformatics Group, Ziegelstr. 7/9, 10117, [2]Institute of Biochemistry, Computational Systems Biochemistry Group, Seestr. 73, 13347 Berlin, Germany and [3]Laboratory of Structural Bioinformatics, Institute of Biology and Molecular Biotechnology, Adam Mickiewicz University, ul. Umultowska 89, 61-614 Poznań, Poland

Received August 15, 2012; Revised October 10, 2012; Accepted October 11, 2012

ABSTRACT

Voronoia4RNA (http://proteinformatics.charite.de/voronoia4rna/) is a structural database storing precalculated atomic volumes, atomic packing densities (PDs) and coordinates of internal cavities for currently 1869 RNAs and RNA–protein complexes. Atomic PDs are a measure for van der Waals interactions. Regions of low PD, containing water-sized internal cavities, refer to local structure flexibility or compressibility. RNA molecules build up the skeleton of large molecular machineries such as ribosomes or form smaller flexible structures such as riboswitches. The wealth of structural data on RNAs and their complexes allows setting up representative data sets and analysis of their structural features. We calculated atomic PDs from atomic volumes determined by the Voronoi cell method and internal cavities analytically by Delaunay triangulation. Reference internal PD values were derived from a non-redundant sub-data set of buried atoms. Comparison of internal PD values shows that RNA is more tightly packed than proteins. Finally, the relation between structure size, resolution and internal PD of the Voronoia4RNA entries is discussed. RNA, protein structures and their complexes can be visualized by the Jmol-based viewer Provi. Variations in PD are depicted by a color code. Internal cavities are represented by their molecular boundaries or schematically as balls.

INTRODUCTION

Atomic packing density (PD) is a measure applied to investigate various aspects of RNA or protein structures. Analysis of PD serves as quality check of tertiary structure models (1), to estimate van der Waals (vdW) forces (2,3) and to detect regions of intrinsic flexibility (4) or compressibility (5,6). In addition, volumetric calculations were implemented to determine and to analyze ligand-binding pockets (7,8). Frequent application and evaluation of PDs for proteins led to the implementation of several applications and databases to make the findings accessible (9–11). However, at the moment no such application or database for RNAs and RNA–protein complexes (referred to as RNA structures) exists.

Regions of low internal PDs often contain water-sized cavities (referred to as internal cavities). The analysis of channels and transporters revealed that internal cavities cluster along the channel's pore or within the hinge regions of transporters. As a result, the internal PD of channels and transporters turned out to be lower than those of water soluble helix bundle proteins (3). The internal PD of membrane proteins, which structures are considered to be more rigid, again turned out to be comparable to those of helix bundle water soluble proteins. The more general outcome of these analyses was that vdW interactions do not energetically compensate for the absence of the hydrophobic effect within the lipid bilayer (4). Moreover, it shows that the combination of analysis of internal PD and internal cavities provides a valuable tool to investigate structure–function relationships of biomolecules.

*To whom correspondence should be addressed. Tel: +49 30 450 524 190; Fax: +49 30 450 524 952; Email: peter.hildebrand@charite.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Initial analysis of PDs of 50 non-redundant RNA structures by means of the Voronoi method shows that these structures are packed more tightly than proteins (12). Analysis of hydrogen bonds and hydrophobic contacts on a topological network representation of RNA structures was again used to investigate and predict flexible sites (13,14). A measure for atomic PD in conjunction with an analysis of hydrogen bonds was used to evaluate flexible sites of the ribosomal 16S RNA identified by ultraviolet-induced photocrosslinking (15). To reveal sites of high PD biochemically, changes to the three-dimensional structure of RNA upon mutation of residues to analogs with higher volumes were investigated (16). A statistical analysis of geometric features was used to analyze the shape and packing of interface contacts in RNA–protein complexes (17,18). However, in contrast to protein structures (10), volumetric methods remain to be applied to the combined analysis of PD and molecular function in RNA structures.

The number of RNA structures deposited in the PDB is steadily increasing (19). Although between 1976 and 2005 only ~700 RNA structures were deposited in the PDB, >1500 entries have been added since then. Crystallization, X-ray and electron microscopy (EM) analysis of ribosomal subunits from both prokaryotes (20–22) and eukaryotes (23–26) revealed the three-dimensional organization of ribosomes and provide a large data pool for further analysis. The different folding motifs of RNA structures have been studied in great detail as reviewed in (27) and stored in several databases (28,29). Atomic level tools have been developed to analyze and classify particular base–base interactions (30) and to assess the quality of RNA tertiary structure models (31,32). However, so far, no tool has been available for detailed analysis and visualization of PDs and internal cavities of RNA structures.

Here, we introduce Voronoia4RNA, a database that stores PDs, and internal cavities from presently 1869 RNA containing PDB structures, resolved at 3.5 Å or better. For the calculation of PD, vdW volumes were derived from the Voronoi cell method (33). It calculates atomic volumes based on the allocation of space among atoms using hyperbolic surfaces. A non-redundant sub-data set of 606 entries was analyzed to calculate reference PDs for different groups of atoms (34). In line with previous analysis (12), we find higher average internal PD values for RNA when compared with proteins. We further discuss how internal PD relates with size and resolution of the structures. The functionality of the website will be described in detail, i.e. how different data sets can be composed based on search terms, compared with reference values, visualized by the Jmol-based molecule viewer Provi and downloaded for further analysis. Thus, Voronoia4RNA aims to help clarify the structure–function relationships of RNA structures by providing data sets for large-scale statistical analysis as well as an integrated molecular viewer for the exploratory analysis of particular structures.

METHODS AND DESCRIPTION

Voronoi cell method

The original Voronoi method (35) allocates the space between points by constructing planes in equidistance of every pair of neighboring points. In the three-dimensional space, the constructed planes intersect and thereby form polyhedral volumes enclosing these points. In the context of RNA or protein molecules, the polyhedral volumes are created around atom positions to estimate the volume occupied by each atom. Furthermore, the planes are constructed at a distance weighted by the atoms vdW radii to take different atom sizes into account.

For Voronoia4RNA, we applied a specified version of the original Voronoi method, the Voronoi cell method, that uses hyperbolic surfaces instead of planes to allocate the atomic volumes (33). First, the solvent-excluded volume of a structure is calculated with a probe radius of 1.4 Å to differentiate between buried and exposed atoms. Then, for each atom two different volumes are calculated. V_{vdW} is the volume assigned to each atom by the Voronoi cell method but only inside the atoms vdW sphere. V_{se} is the remaining solvent excluded volume assigned to each atom. From these two volumes, the PD for every atom is calculated as $PD = V_{vdW}/(V_{vdW} + V_{se})$.

The absolute value of the PD depends directly on the used set of atom radii. Voronoia4RNA uses the ProtOr (36) atom radius set for protein atoms and NucProt (12) set for RNA atoms that were both determined analytically from reference structures. Atom types from RNA and proteins are grouped according to their chemical nature, valence and number of attached hydrogen atoms. To determine the vdW radii of these atomic groups, the contacts of each atom group were analyzed and an average radius was calculated (12,36). Internal cavities are determined analytically from the structures by a Delaunay triangulation (37). Internal cavities are defined as a buried space within a structure that is big enough to accommodate at least a 1.4 Å radius water-sized probe.

Database description

For Voronoia4RNA, only PDB entries resolved by X-ray or EM at 3.5 Å or better, or by nuclear magnetic resonance (NMR) containing at least one RNA chain were considered. At the moment Voronoia4RNA is composed of 1340 X-ray, 524 NMR, 4 fiber diffraction and 1 EM structure summing up to 1869 entries; 947 of these structures contain one or more proteins. Voronoia4RNA can be searched for PDB entry codes or for keywords contained in the header of the PDB file. Results of a search can be visualized on-line (see below) or downloaded for further analysis. To calculate mean PD values for each atom group, we analyzed a reference data set of RNA structures derived from the database of the BGSU RNA Structural Bioinformatics Group. They use a length-dependent redundancy criteria (38) to compile non-redundant data sets of RNA structures (http://rna.bgsu.edu/nrlist/). The 2012-05-05 release contains 621 structures with a resolution of at least 3.5 Å. After removing 15 synthetic RNA constructs, our reference data set contains

606 entries for which we calculated reference internal PDs. We solely investigate the internal packing of the structures and not the packing with the solvent, thus we exclude non-buried atoms.

Data processing

In the first step, the input file, which is ideally in the standard PDB format, is preprocessed to correct for common aberrations from the PDB standard, such as misaligned columns or uncommon atom names. Only internal (buried) water molecules are considered for the calculation of PDs. All other water molecules are ignored. Non-water heteroatoms are only considered if they are in close contact (at least 2.8 Å) to the structure. From NMR structures, the first model is selected for the calculation of PDs and internal cavities. Split PDB entries are generally merged. In the second step, V_{vdW} and V_{se} are calculated by the Voronoi cell method and stored in the volume file, which is a PDB file extended by the atomic PD values. Additionally, an extended volume file is created that also lists z-scores for each atom and for the complete structure.

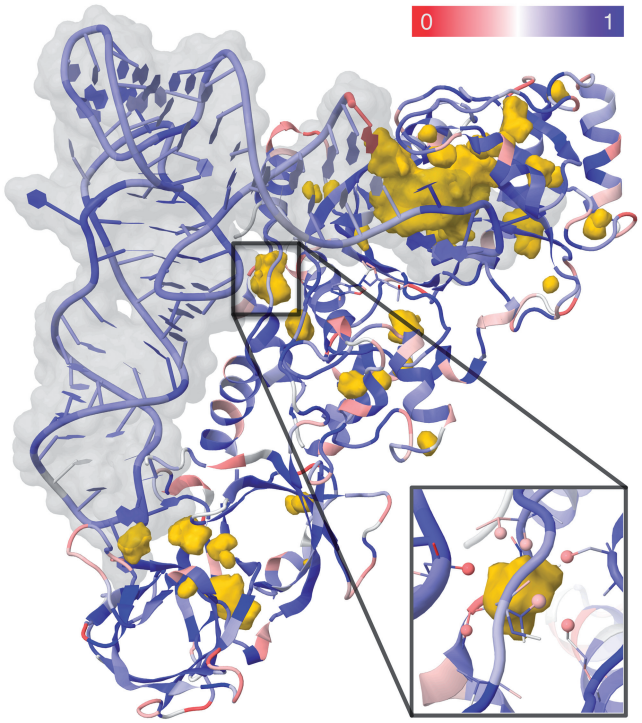

Figure 1. PD and internal cavities of a sample structure. Visualization of the *Escherichia coli* glutaminyl-tRNA synthetase in complex with tRNA2Gln and ATP (PDB entry: 1GTS). The structure is colored according to the PD values of the C-α (proteins) and the phosphate atoms (RNA) by the color-scale shown in the top right corner. Regions of low PD (colored red) are prevalent within the protein surface region and regions of high PD (dark blue) within the RNA part. Internal cavities represented by their molecular boundaries (gold) cluster within the protein and the protein–tRNA interface. The large cavity in the upper right region of the complex marks the glutamine-binding site at the 3′-end of the tRNA. To distinguish the RNA from the protein part, the tRNA surface is shown in translucent gray. The inset shows the neighboring residues/nucleotides (wire-frame) with the atoms that delineate the cavity (gold) highlighted as small balls and colored according to their PD values.

Here, the z-score is the normalized difference between the internal PD value and the according reference internal PD value.

Visualization

PDs and internal cavities can be visualized on-line with Provi in the context of their three-dimensional structure (Figure 1). Provi is a Jmol-(http://www.jmol.org/) based web application that visualizes macro-molecular structures and related data. Atoms are colored according to their PD value on a color scale from red (value 0.0) over white (0.5) to blue (1.0). An internal cavity is represented as a sphere, placed at the geometric center of the atoms that delineate a cavity, with diameter approximating the cavity sizes. Additionally, internal cavities can be represented by their molecular boundaries.

DISCUSSION

The large number of diverse RNA structures, including large molecular machineries such as ribosomes or smaller flexible structures such as riboswitches, now allows setting up and analyzing representative data sets. Here, we introduced Voronoia4RNA, a database containing data on atomic PDs and internal cavities of 1869 RNAs and RNA–protein complexes. Our analysis of a reverence sub-data set of 606 structures reveals higher average internal PD values of RNA (0.79, SD: 0.09) than of proteins (0.71, SD: 0.11). This confirms results of a previous analysis of a smaller data set of 50 RNA structures showing that RNA is packed more tightly than proteins (12).

Voronoia4RNA contains entries of different sizes in the order of 10^2–10^5 atoms with resolutions ranging from 0.6 to 3.5 Å. To investigate whether there is a relation between size, resolution and internal PDs, size (Figure 2) and resolution (Figure 3) were plotted against the reference internal PD values. The former plot indicates that there

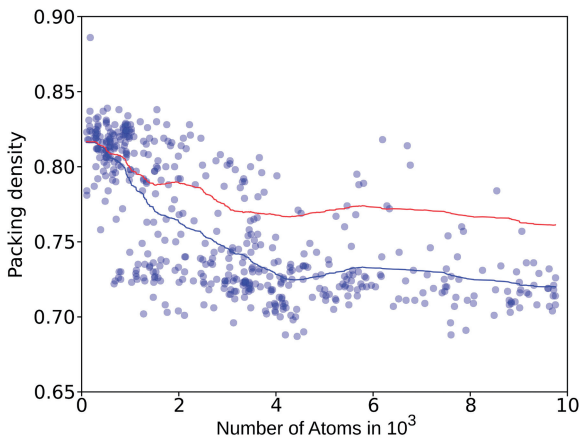

Figure 2. Average internal PD values of RNA structures (blue dots) plotted against the number of atoms. The average internal PD of the structures of the reference data set decreases up to a size of ~3000 atoms, as denoted by the running average (blue line) drawn on top of the raw data. When only the RNA part of the structures is considered (red line), the decrease in internal PD rests on a higher plateau due to the overall higher internal PD of RNA.

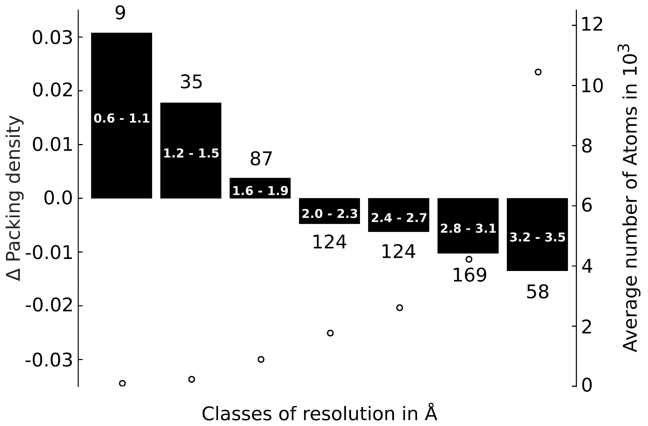

Figure 3. Deviation of internal PD from the overall mean value for different classes of resolution. Structures obtained with a resolution better than 2 Å show internal PD values higher as the average. Those with a resolution of 2 Å and more show internal PD values below the average. The number of representatives of a class is denoted by the black number at the tip of the bar. The resolution of the class is denoted by the white numbers within the bar. The average structure size (in number of atoms) is depicted by a black circle.

is a relation between structure size and internal PD. Structures with <3000 atoms show considerably higher internal PD values than the structures with more atoms (Figure 2). This is particularly true for RNA–protein complexes (Figure 2, blue line), while the RNA part alone shows a weaker relation between size and internal PD (Figure 2, red line). The latter plot shows that structures resolved at higher resolution also have a tendency for higher internal PD values (Figure 3). To elucidate whether there is a correlation between structure size and resolution, these two variables were also plotted against each other. We find that the average structure size correlates with the resolution (Figure 3, black circles).

The observed relation between structure size, resolution and PD can arise from several causes. For proteins, it has been described that internal cavities and smaller internal packing defects only occur in structures with a minimum size (10). As a rule, protein structures >150 residues contain internal cavities. In analogy, small RNA structures may not contain internal cavities and may also be tightly packed. Next, larger assemblies contain many intermolecular surfaces at which internal cavities can be found (39). In low-resolution structures of large assemblies, such as the ribosome, not all internal water molecules are resolved. This results in more empty internal cavities and lower average internal PDs. It remains elusive, to which extent these packing defects have a functional role, i.e. contribute to internal flexibilities or are simply an effect of the lower structure resolution. Taken together, our initial analysis of the data stored in Voronoia4RNA provides promising starting points for more detailed explorations of PDs of RNA structures.

FUNDING

Deutsche Forschungsgemeinschaft [DFG HI 1502/1-1, SFB 740] and European Research Council [ERC Advanced Grant TUDOR]. Funding for open access charge: SFB 740/B6.

Conflict of interest statement. None declared.

REFERENCES

1. Pontius,J., Richelle,J. and Wodak,S.J. (1996) Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J. Mol. Biol.*, **264**, 121–136.
2. Eilers,M., Shekar,S.C., Shieh,T., Smith,S.O. and Fleming,P.J. (2000) Internal packing of helical membrane proteins. *Proc. Natl Acad. Sci. USA*, **97**, 5796–5801.
3. Hildebrand,P.W., Rother,K., Goede,A., Preissner,R. and Frömmel,C. (2005) Molecular packing and packing defects in helical membrane proteins. *Biophys. J.*, **88**, 1970–1977.
4. Hildebrand,P.W., Günther,S., Goede,A., Forrest,L., Frömmel,C. and Preissner,R. (2008) Hydrogen-bonding and packing features of membrane proteins: functional implications. *Biophys. J.*, **94**, 1945–1953.
5. Dadarlat,V.M. and Post,C.B. (2001) Insights into protein compressibility from molecular dynamics simulations. *J. Phys. Chem. B*, **105**, 715–724.
6. Dadarlat,V.M. and Post,C.B. (2006) Decomposition of protein experimental compressibility into intrinsic and hydration shell contributions. *Biophys. J.*, **91**, 4544–4554.
7. Liang,J., Edelsbrunner,H. and Woodward,C. (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.*, **7**, 1884–1897.
8. Chen,B.Y. and Honig,B. (2010) VASP: a volumetric analysis of surface properties yields insights into protein–ligand binding specificity. *PLoS Comput. Biol.*, **6**, e1000881.
9. Tsai,J. and Gerstein,M. (2002) Calculations of protein volumes: sensitivity analysis and parameter database. *Bioinformatics*, **18**, 985–995.
10. Rother,K., Preissner,R., Goede,A. and Frommel,C. (2003) Inhomogeneous molecular density: reference packing densities and distribution of cavities within proteins. *Bioinformatics*, **19**, 2112–2121.
11. Rother,K., Hildebrand,P.W., Goede,A., Gruening,B. and Preissner,R. (2009) Voronoia: analyzing packing in protein structures. *Nucleic Acids Res.*, **37**, D393–D395.
12. Voss,N.R. and Gerstein,M. (2005) Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly. *J. Mol. Biol.*, **346**, 477–492.
13. Fulle,S. and Gohlke,H. (2008) Analyzing the flexibility of RNA structures by constraint counting. *Biophys. J.*, **94**, 4202–4219.
14. Fulle,S. and Gohlke,H. (2009) Constraint counting on RNA structures: linking flexibility and function. *Methods*, **49**, 181–188.
15. Huggins,W., Ghosh,S.K. and Wollenzien,P. (2009) Hydrogen bonding and packing density are factors most strongly connected to limiting sites of high flexibility in the 16S rRNA in the 30S ribosome. *BMC Struct. Biol.*, **9**, 49.
16. Schwans,J.P., Li,N.-S. and Piccirilli,J.A. (2004) A packing-density metric for exploring the interior of folded RNA molecules. *Angew. Chem. Int. Ed Engl.*, **43**, 3033–3037.
17. Bahadur,R.P., Zacharias,M. and Janin,J. (2008) Dissecting protein–RNA recognition sites. *Nucleic Acids Res.*, **36**, 2705–2716.
18. Rawat,N. and Biswas,P. (2011) Shape, flexibility and packing of proteins and nucleic acids in complexes. *Phys. Chem. Chem. Phys.*, **13**, 9632–9643.
19. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
20. Wimberly,B.T., Brodersen,D.E., Clemons,W.M., Morgan-Warren,R.J., Carter,A.P., Vonrhein,C., Hartsch,T. and Ramakrishnan,V. (2000) Structure of the 30S ribosomal subunit. *Nature*, **407**, 327–339.
21. Ban,N. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.

22. Gabashvili,I.S., Agrawal,R.K., Spahn,C.M., Grassucci,R.A., Svergun,D.I., Frank,J. and Penczek,P. (2000) Solution structure of the *E. coli* 70S ribosome at 11.5 Å resolution. *Cell*, **100**, 537–549.
23. Ben-Shem,A., Jenner,L., Yusupova,G. and Yusupov,M. (2010) Crystal structure of the eukaryotic ribosome. *Science*, **330**, 1203–1209.
24. Ben-shem,A., Melnikov,S., Jenner,L. and Yusupova,G. (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science*, **334**, 1524–1529.
25. Spahn,C., Beckmann,R. and Eswar,N. (2001) Structure of the 80S ribosome from *Saccharomyces cerevisiae*—tRNA-ribosome and subunit–subunit interactions. *Cell*, **107**, 373–386.
26. Ratje,A.H., Loerke,J., Mikolajka,A., Brünner,M., Hildebrand,P.W., Starosta,A.L., Dönhöfer,A., Connell,S.R., Fucini,P., Mielke,T. *et al.* (2010) Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. *Nature*, **468**, 713–716.
27. Holbrook,S.R. (2005) RNA structure: the long and the short of it. *Curr. Opin. Struct. Biol.*, **15**, 302–308.
28. Vanegas,P.L., Hudson,G.A., Davis,A.R., Kelly,S.C., Kirkpatrick,C.C. and Znosko,B.M. (2012) RNA CoSSMos: characterization of secondary structure motifs—a searchable database of secondary structure motifs in RNA three-dimensional structures. *Nucleic Acids Res.*, **40**, D439–D444.
29. Bindewald,E., Hayes,R., Yingling,Y.G., Kasprzak,W. and Shapiro,B.A. (2008) RNAJunction: a database of RNA junctions and kissing loops for three-dimensional structural analysis and nanodesign. *Nucleic Acids Res.*, **36**, D392–D397.
30. Petrov,A.I., Zirbel,C.L. and Leontis,N.B. (2011) WebFR3D—a server for finding, aligning and analyzing recurrent RNA 3D motifs. *Nucleic Acids Res.*, **39**, W50–W55.
31. Davis,I.W., Leaver-Fay,A., Chen,V.B., Block,J.N., Kapral,G.J., Wang,X., Murray,L.W., Arendall,W.B., Snoeyink,J., Richardson,J.S. *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375–W383.
32. Parisien,M. (2009) New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA*, **15**, 1875–1885.
33. Goede,A. and Preissner,R. (1996) Voronoi cell: new method for allocation of space among atoms: elimination of avoidable errors in calculation of atomic volume and density. *J. Comput. Chem.*, **18**, 1113–1123.
34. Tsai,J., Voss,N. and Gerstein,M. (2001) Determining the minimum number of types necessary to represent the sizes of protein atoms. *Bioinformatics*, **17**, 949–956.
35. Voronoi,G. (1908) Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs. *Journal für die reine und angewandte Mathematik*, **1908**, 198–287.
36. Tsai,J., Taylor,R., Chothia,C. and Gerstein,M. (1999) The packing density in proteins: standard radii and volumes. *J. Mol. Biol.*, **290**, 253–266.
37. Delaunay,B. (1934) Sur la sphere vide. *Bull. Acad. Sci. USSR*, **6**, 793–800.
38. Leontis,N.B. and Zirbel,C.L. (2012) Nonredundant 3D Structure Datasets for RNA Knowledge Extraction and Benchmarking. In: Leontis,N. and Westhof,E. (eds), *RNA 3D structure analysis and prediction*. Springer, Berlin, pp. 281–298.
39. Janin,J. (1999) Wet and dry interfaces: the role of solvent in protein–protein and protein–DNA recognition. *Structure*, **7**, R277–R279.